

**Proyecto Júpiter – Propuesta de IA  
generativa  
DETECCIÓN DE FAKE NEWS  
Pontiapp**

**Pontia - Octubre 2024**

Lucía Cámara Chamorro  
Amparo Díaz Román  
María Belén Libonati Cattai  
Carolina Méndez

## PROPUESTA DE IA GENERATIVA

### La IA generativa y su potencialidad para Pontiapp

La IA generativa está en pleno auge actualmente, siendo cada vez más las organizaciones que la utilizan para mejorar la creación de contenido en distintas áreas. Un ejemplo de ello es su aplicación en el marketing, donde permite a las empresas ofrecer experiencias más personalizadas y relevantes a los clientes (Kshetri, Dwivedi, Davenport y Panteli, 2024). Grandes empresas como Meta y OpenAI ya utilizan herramientas de IA generativa para optimizar tiempo y recursos y reducir costes (Lee, Lee, Jeong y Kim, 2024). Pontiapp también podría beneficiarse de la implementación de estas herramientas. Entre los posibles usos que Pontiapp podría hacer de la IA generativa se encuentran la creación de listas negras de fuentes de publicación, la corrección de noticias falsas y el filtrado inteligente y resumen personalizado de noticias.

La creación de listas negras actualizadas con relativa periodicidad permitiría filtrar de forma automática aquellas noticias con mayor probabilidad de ser falsas, dependiendo de la credibilidad atribuida a su fuente de publicación. Este filtro rápido ayudaría a reducir los costes derivados de la implementación continua de modelos de aprendizaje más complejos para la discriminación entre noticias reales y falsas. También reduciría la probabilidad de difusión de las noticias falsas, evitando la propagación de bulos entre la sociedad, con consecuencias indeseadas tanto para la sociedad en sí misma como para el prestigio de la empresa.

A lo mencionado anteriormente también ayudaría la aplicación de la IA generativa para la corrección de noticias falsas. Cuanto más rápida fuese la actuación de Pontiapp para detectar y corregir las noticias falsas, mayor sería la confianza que depositarían los clientes en esta empresa. Ningún cliente desea utilizar una app de mensajería de cuestionable credibilidad que no se preocupa de contrastar sus fuentes. En cambio, la percepción de que la empresa implementa procedimientos adecuados, tanto para la detección de noticias falsas como para la rápida corrección de estas noticias, aumentaría considerablemente el grado de satisfacción de los clientes.

Otra alternativa que sin duda potenciaría el grado de satisfacción de los clientes sería poder estar al día de las noticias más novedosas y relevantes, sin necesidad de tener que invertir demasiado tiempo en esta tarea. Teniendo en cuenta el ritmo de vida actual, parece bastante claro las limitaciones en cuanto al tiempo del que disponen los clientes para acceder a noticias actualizadas y de interés. Por ello, cualquier empresa que les facilitase esta cuestión sería bien acogida en el mercado. Este sería precisamente el objetivo de aplicar la IA generativa al filtrado inteligente y resumen personalizado de noticias según los intereses de los clientes.

Posibles nombres comerciales para las potenciales aplicaciones sugeridas de la IA generativa podrían ser: a) *BlackList*, para la creación de listas negras; b) *FactNews*, para la corrección de noticias falsas; y c) *TailorNews*, para el filtrado inteligente y el resumen personalizado de noticias. En general, consideramos que las tres aplicaciones propuestas se traducirían en una mayor competitividad de la empresa en el mercado y que los tres nombres comerciales elegidos resultarían atractivos para los clientes. Sin embargo, debido a las necesidades actuales de la sociedad en la que vivimos, creemos que el filtrado inteligente y el resumen personalizado de noticias es la propuesta que más podrían demandar los clientes. Por ello, es TailorNews la primera propuesta por la que se apuesta en este proyecto, y es su proceso de implementación el que se explica de forma más o menos detallada a continuación.

### **Procedimiento de implementación de TailorNews**

El filtrado inteligente y resumen personalizado de noticias atendiendo a los intereses y necesidades de los clientes se llevaría a cabo mediante tres fases principales:

- 1) Preselección de las noticias según su popularidad.
- 2) Selección de las noticias según la credibilidad de su fuente de publicación.
- 3) Traducción y/o resumen de las noticias seleccionadas.

#### ***Preselección de las noticias según su popularidad***

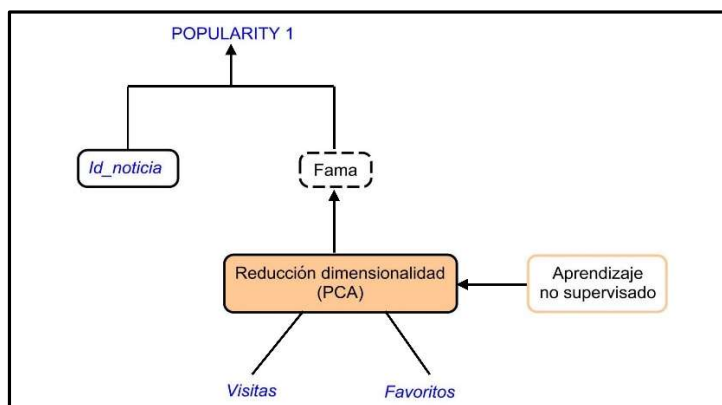
Las noticias serían preseleccionadas según una puntuación total en popularidad, calculada a partir de tres indicadores o puntuaciones distintas de popularidad:

- 1) Fama o puntuación en Popularidad 1: determinada por el número de visitas y la puntuación como favorita recibida por cada noticia.
- 2) Origen o puntuación en Popularidad 2: determinada por el número de visitas recibidas por las noticias según su fuente, país e idioma de publicación.
- 3) Categoría o puntuación en Popularidad 3: determinada por el número de visitas recibidas por las noticias según el contenido y tipo de noticia.

Para el cálculo de las distintas puntuaciones se utilizarían diferentes estrategias de aprendizaje automático, optando por dos modelos de *machine learning* clásicos para calcular la Popularidad 1 y la Popularidad 2 y por un modelo *transformer (language large model [LLM], o deep learning)* para calcular la Popularidad 3:

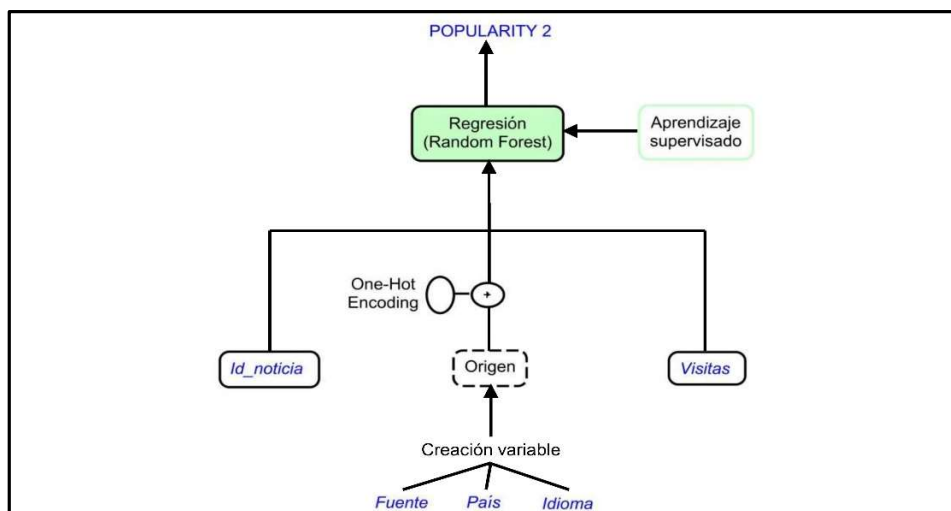
**Popularidad 1.** El número de visitas y las puntuaciones como favoritas se reducirían a una sola dimensión (fama) mediante un análisis de componentes principales (*principal component analysis* o PCA). Se optaría por esta estrategia de aprendizaje no supervisado por dos motivos principales. El primero, la disponibilidad de datos de entrada (Visitas y Favoritos), pero no de datos de salida, lo que hace más adecuado usar una estrategia de aprendizaje no supervisado que una de aprendizaje

supervisado. El segundo, el deseo de evitar el mayor esfuerzo y las complicaciones que podría conllevar el usar otras estrategias al calcular la puntuación en Fama, como su cálculo manual (implicaría crear la variable fama teniendo que asignar un peso a cada una de las otras dos variables), o el uso de un algoritmo de *clustering* (supondría tener que decidir el número de *clusters* o grupos en los que dividir a las noticias por popularidad). Con la variable Fama y la variable *Id\_noticia* se obtendría manualmente en Python la lista de noticias ordenadas según su puntuación en fama (Figura 1).



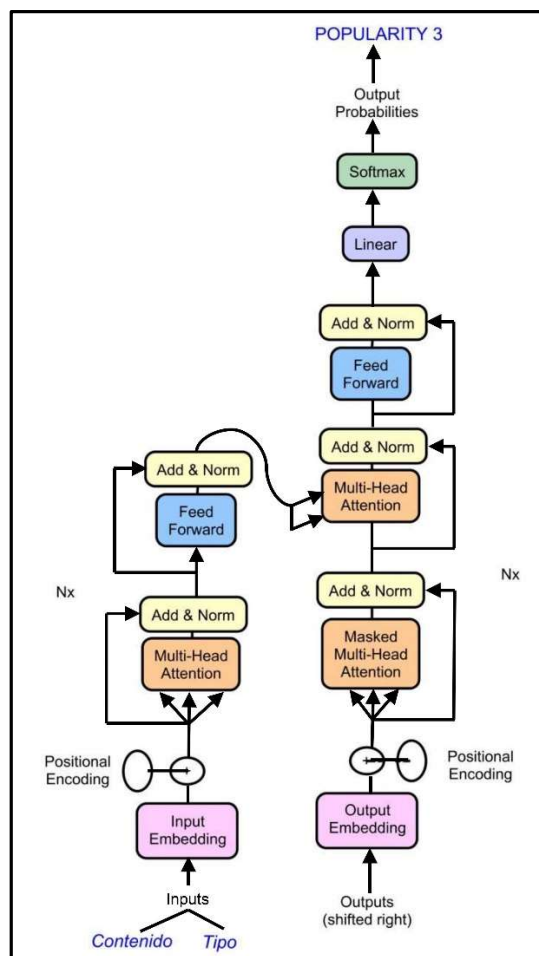
**Figura 1.** Proceso de obtención de las noticias ordenadas según su puntuación en fama.

**Popularidad 2.** Con las variables cualitativas Fuente, País e Idioma se crearía manualmente en Python la variable Origen. Esta variable se convertiría en formato numérico con *One-Hot Encoding*, optándose por este método de codificación binaria por tratarse de una variable sin un orden inherente entre las categorías. Posteriormente, con las variables Origen, *Id\_noticia* y Visitas se obtendría el número de visitas que recibiría cada noticia según su origen utilizando un algoritmo de aprendizaje supervisado: *Random Forest Regression* (Figura 2). La selección de este algoritmo sería debida a la disponibilidad de datos de entrada y salida en este caso (aprendizaje supervisado frente a no supervisado) y a que este algoritmo permite trabajar a la vez con datos cualitativos y cuantitativos (Origen y Visitas) y admite relaciones no lineales entre las variables. Al final se obtendría la lista de noticias ordenadas según el número de visitas previstas por su origen. Cabe mencionar que el tener en cuenta la variable Visitas en lugar de la de Favoritos para calcular esta popularidad se debe a que el número de visitas recoge mejor lo que se quiere valorar. A mayor número de visitas recibidas, mayor interés ha despertado de primeras una noticia, independientemente de que después haya podido gustar más o menos.



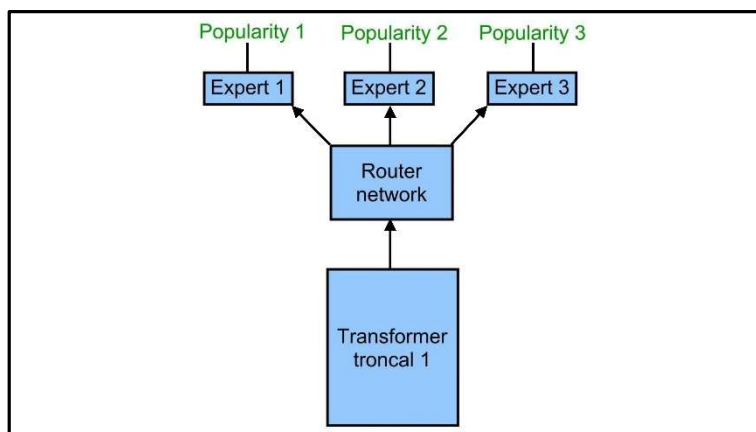
**Figura 2.** Proceso de obtención de las noticias ordenadas según las visitas previstas por su origen.

**Popularidad 3.** Con un modelo transformer (*encoder-decoder*) y las variables Contenido y Tipo como inputs, se obtendría la lista de noticias ordenadas según el número de visitas previstas para cada una de ellas por su categoría. En la Figura 3 se muestra la arquitectura transformer, que incorpora tanto los *embeddings* requeridos para vectorización de textos y mantener el orden de las palabras, como las capas de atención necesarias para que el modelo entienda la relación existente entre las palabras.



**Figura 3.** Arquitectura Transformer para la obtención de las noticias ordenadas según las visitas previstas por su categoría.

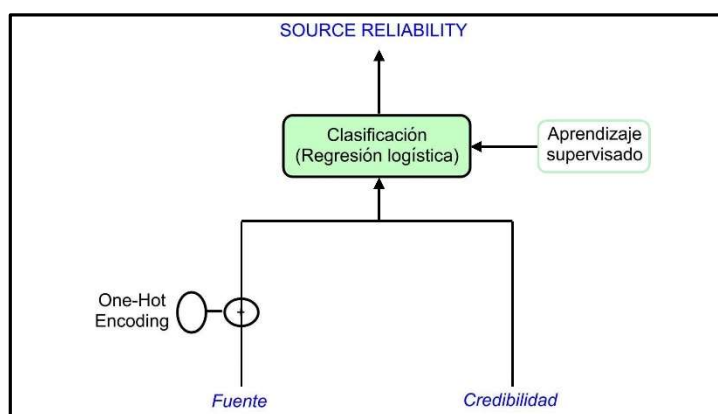
Una vez obtenidas las tres puntuaciones en popularidad, se combinarían con la técnica *Mixture of Experts* (MoE), utilizando un transformer troncal, un *router* y tres redes neuronales o expertos (uno por puntuación; Figura 4). El resultado sería la popularidad total, puntuación que serviría para obtener la lista final de noticias ordenadas por popularidad de la cual se preseleccionarían las 40 primeras. Estas 40 noticias serían con las que se trabajaría en la siguiente fase de la propuesta. El uso de la técnica MoE permitiría ahorrar recursos, posibilitando también asignar distinto peso o importancia a las distintas puntuaciones en popularidad para el cálculo de la puntuación total (por medio del router), o incluso prescindir de alguna de ellas según las preferencias de la empresa en cada momento.



**Figura 4.** Mixture of Experts para la obtención de la puntuación total en popularidad.

### ***Selección de las noticias según la credibilidad de su fuente de publicación***

Las variables Fuente y Credibilidad (noticia real o falsa) se utilizarían para obtener una lista de fuentes ordenadas según la probabilidad de publicar noticias reales o falsas. El proceso pasaría por la codificación numérica (One-Hot Encoding) de la variable Fuente (la variable Credibilidad no porque ya se encuentra en formato numérico 1/0) y el uso de un algoritmo de clasificación o regresión logística (aprendizaje supervisado; Figura 5). Las 40 noticias preseleccionadas serían ordenadas manualmente con Python, atendiendo a su posición en la lista ordenada de credibilidad, y se seleccionarían las 10 primeras.

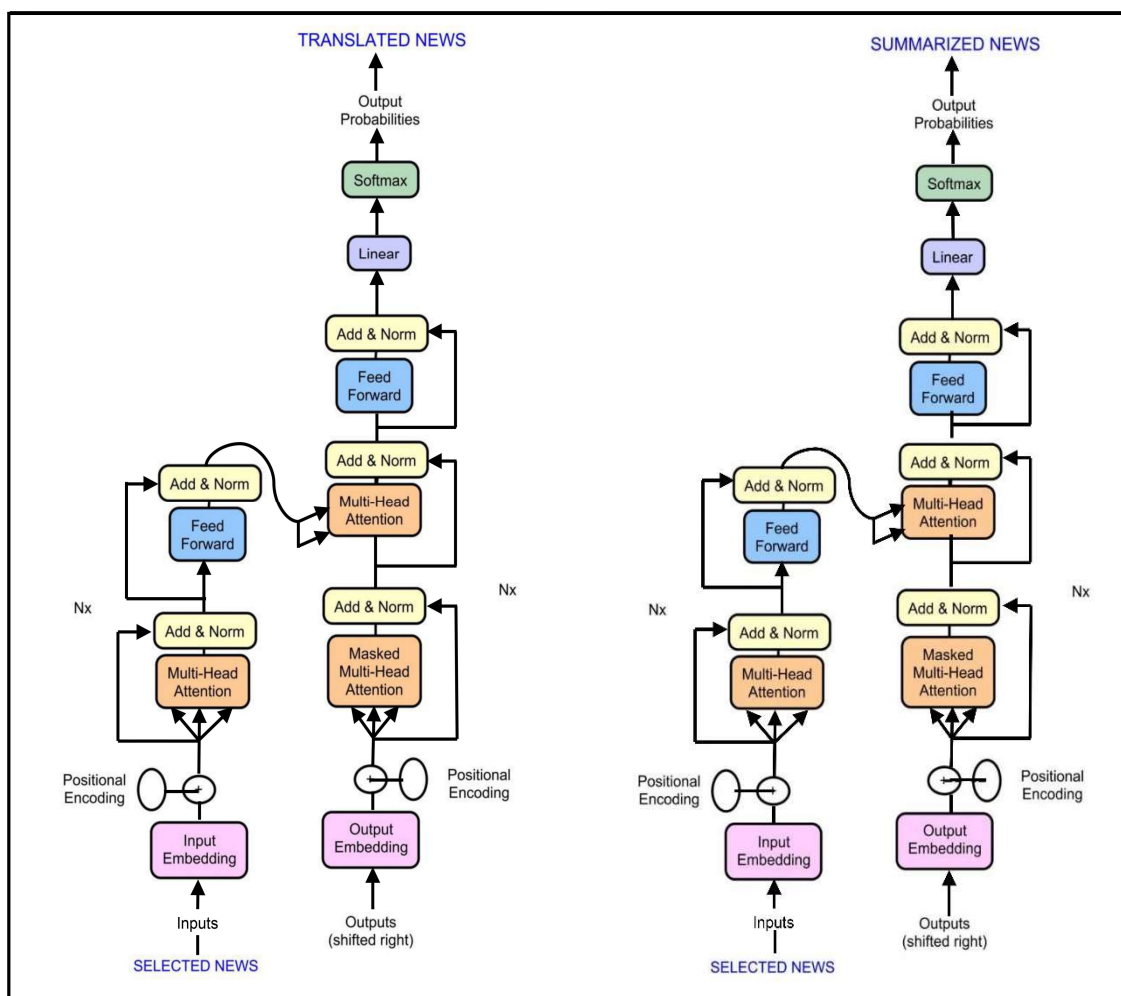


**Figura 5.** Proceso de obtención de las fuentes ordenadas según probabilidad de publicar noticias reales o falsas.

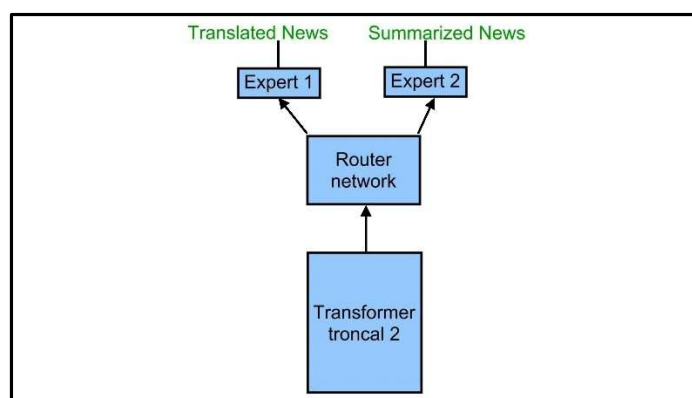
### ***Traducción y/o resumen de las noticias seleccionadas***

Las 10 noticias finalmente seleccionadas serían traducidas en caso necesario utilizando un modelo transformer y resumidas utilizando otro modelo transformer (Figura 6). Nuevamente se emplearía la técnica MoE, con otro transformer troncal y dos expertos en esta ocasión, para reducir

costes y porque no siempre será necesaria la traducción previa de las noticias (Figura 7). El producto final de este proceso, mostrado resumido en la Figura 8, sería ya TailorNews.

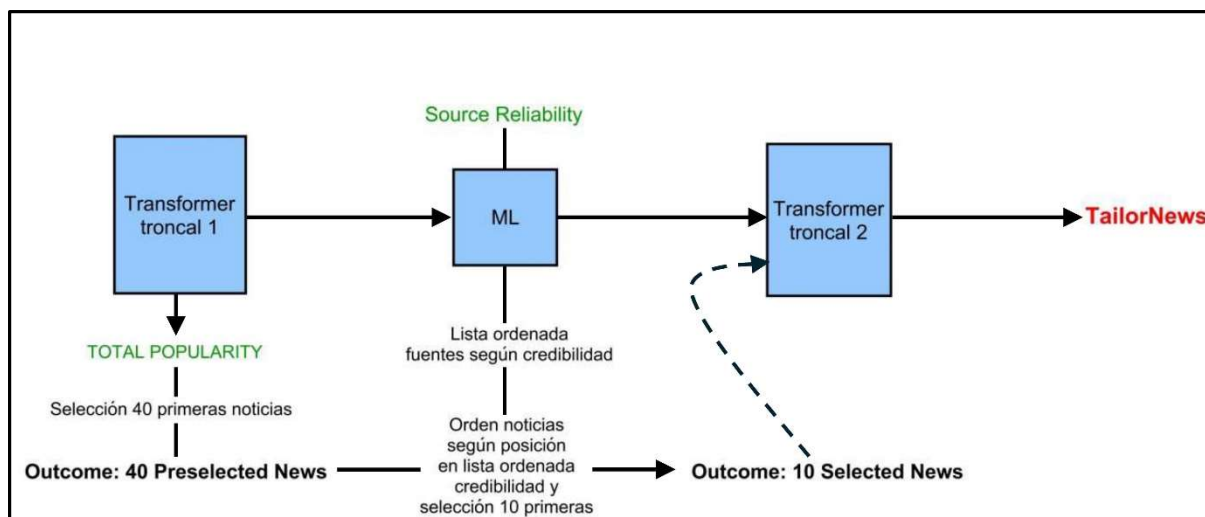


**Figura 6.** Transformers para la traducción y/o resumen de las noticias seleccionadas.



**Figura 7.** Mixture of Experts para la traducción y/o resumen de las noticias seleccionadas.





**Figura 8.** Proceso de desarrollo de TailorNews.

### Evaluación de resultados

Para evaluar el éxito de la implementación de la IA generativa en Pontiapp con TailorNews se podrían utilizar varios indicadores, evaluados antes y después de su implementación, como los siguientes: el número de clientes o usuarios que utilizan la aplicación, su frecuencia de uso frente a otras aplicaciones con fines similares y su grado de satisfacción. La comparación pre y post de estos indicadores constituiría una buena medida de la adaptación de Pontiapp al mercado y de su competitividad dentro del mismo. Si bien sería necesario considerar también otros indicadores como los costes y beneficios que ha supuesto el desarrollo de TailorNews (ej., capacitación de personal, adquisición de infraestructura o software especializado, etc.), los cuales serían clave para evaluar la verdadera rentabilidad de la implementación de la IA generativa en Pontiapp.

### Posibles riesgos

La implementación de herramientas de IA generativa no está exenta de algunos riesgos que toda empresa que apueste por ello debe tener en cuenta. Entre estos riesgos se encuentran: a) los propios costes que supone llevar a cabo cualquier proyecto que integre IA generativa, pudiendo ser precisa una considerable inversión en recursos humanos y tecnológicos; b) los sesgos algorítmicos en los que podría incurrirse por la falta de exactitud o representatividad de los datos, la estructura de los modelos o la forma de implementar la IA generativa; y c) la sobrepersonalización o excesiva adaptación de la información a las preferencias de los clientes, pudiendo dar lugar a la difusión de información limitada o falsa (Akter et al., 2023; Raman et al., 2024; Wolfe y Mitra, 2024). La programación de reuniones periódicas para la discusión de los avances del proyecto podría ser una forma adecuada de gestionar estos posibles riesgos.

## Conclusiones

Pontiapp podría beneficiarse considerablemente de la implementación de herramientas de IA generativa para distintas tareas, como el filtrado inteligente y resumen personalizado de noticias. La apuesta por TailorNews supondría satisfacer muchas necesidades actuales de los clientes o usuarios cuando desean estar actualizados, aumentando su grado de satisfacción y contribuyendo a mejorar de forma exponencial y duradera la competitividad de la empresa en el mercado. La evaluación y supervisión periódica de los resultados y los posibles riesgos e imprevistos que pudieran surgir facilitaría el avance del proyecto y la consecución de los objetivos. Así, TailorNews podría constituir un buen punto de partida para la implementación de herramientas de IA generativa en Pontiapp, situándola al mismo nivel que grandes empresas internacionales de reconocido prestigio.

## Referencias

- Akter, S., Sultana, S., Mariani, M., Wamba, S. S., Spanaki, K. y Dwivedi, Y. K. (2023). Advancing algorithmic bias management capabilities in AI-driven marketing analytics research. *Industrial Marketing Management*, 114, 243-261. doi:10.1016/j.indmarman.2023.08.013
- Kshetri, N., Dwivedi, Y. K., Davenport, T. H. y Panteli, N. (2024). Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda. *International Journal of Information Management*, 75, 102716. doi:10.1016/j.ijinfomgt.2023.102716
- Lee, G. H., Lee, K. J., Jeong, B. y Kim, T. (2024). Developing personalized marketing service using generative AI. *IEEE Access*, 12, 22394-22402. doi: 10.1109/ACCESS.2024.3361946.
- Raman, R., Nair, V. K., Nedungadi, P., Sahu, A. K., Kowalski, R., Ramanathan, S. y Achuthan, K. (2024). Fake news research trends, linkages to generative artificial intelligence, and sustainable development goals. *Heliyon*, 10, e24727. doi:10.1016/j.heliyon.2024.e24727
- Wolfe, R. y Mitra, T. (2024). The impact and opportunities of generative AI in fact-checking. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, 1531-1543. doi:10.1145/3630106.3658987